DIFFERENCES IN GLOTTAL STOP PERCEPTION BETWEEN ENGLISH AND JAPANESE LISTENERS

Yasuaki Shinohara, Mark Huckvale & Michael Ashby

Department of Speech, Hearing and Phonetic Sciences, University College London, UK
y.shinohara@ucl.ac.uk; m.huckvale@ucl.ac.uk; m.ashby@ucl.ac.uk

ABSTRACT

This study examines whether English L1 and Japanese L2 listeners differ in the way they perceive the glottal stop as a signal to a phonological contrast in English. Glottal stops are often used by native English speakers as an allophone of /t/, including intervocalic environments, while this variation is not found in Japanese. Thus, the different L1 experience of the listeners may cause differences in their perceptual sensitivity to the glottal stop for a lexical contrast in English. In this experiment synthetic continua were constructed between 'bear' [beə] and 'better' [beʔə] using acoustic dimensions of fundamental frequency, voicing amplitude and diplophonia. Identification task results showed English and Japanese listeners differed in terms of how they used the acoustic cues of the glottal gesture for the contrast. Interestingly, Japanese listeners were more sensitive to the amplitude dip than English listeners, suggesting they were more sensitive to the acoustic properties of the stimuli.

Keywords: speech perception, glottal stop, cross-language differences

1. INTRODUCTION

When it comes to phonetic contrasts, there are well established perceptual differences across languages. For example, the way in which voice onset time (VOT) is used to distinguish voiced and voiceless plosives is well known to differ across languages. The perceptual boundary for Spanish listeners is at a smaller VOT than for English listeners because of differences in the laryngeal gestures found in Spanish and English plosive production [13, 14]. These differences between languages can also impact how a listener perceives a second language, since the cues learned for contrasts in L1 may not match those required for the L2. For example, the ability of Japanese listeners to distinguish English /l/ and /r/ is affected by the fact that this contrast is not made in Japanese. This leads to a perceptual insensitivity to the appropriate third formant (F3) cue used by English listeners to detect the difference [10]. A better understanding of these cross-language effects could be useful in teaching and evaluating L2 learners, as well as contributing to models of L1 and L2 language acquisition.

To date, there has been little work on the cross-language perception of the glottal stop. Since there are clear differences between how the glottal stop is used to signal contrasts in English and Japanese, it is an interesting candidate for a new investigation of cross-language perception.

In English Received Pronunciation (RP), the glottal stop is an allophone of particular phonemes in some environments. The voiceless plosives /p, t, k/ and the affricate /tʃ/ in post-vocalic or post-sonorant position are often reinforced with a glottal stop before a pause or before a subsequent consonant [5, 6, 16]. The complete replacement of voiceless plosives with glottal stop is also found in RP, but the most common substitution is for /t/. The context for the occurrence of the glottal allophone is in syllable-final position preceding an obstruent and in word final position preceding an obstruent, approximant or pause [16].

In some regional accents of English, glottal replacement may occur in a wider range of contexts. London-English speakers produce significantly more glottal stops pre-vocalically than speakers of other accent varieties in England [7]. Likewise, glottal replacement for an intervocalic /t/ is a well established feature of London speech [1, 6].

In Japanese, although the glottal stop also functions as an allophone, it appears in very few environments. Some linguists claim that a glottal stop can be a variant of the first part of geminate stops, known as the moraic obstruent /Q/ [9, 17]. An utterance-final instance of the obstruent /Q/ can be realized as a long glottal stop in exclamations or interjections, as in *hey* [koɾaʔː] and *Oh* [aʔː] [3, 15, 17, 18]. A single glottal stop, however, never occurs between vowels as a consonantal allophone. Thus, it is expected that Japanese listeners will differ in their sensitivity to the use of the glottal stop as a signal for a phonological contrast in inter-vocalic position compared to English listeners.

The present study examined how L1 English listeners from London differed from L2 English listeners from Japan in terms of their ability to label synthetic speech continua expressing the absence/presence of an inter-vocalic glottal stop. The task was identification of the contrast between the English words *bear* and *better* using acoustic cues based on a dip in voice fundamental frequency (F0), a dip in voicing amplitude (Ampl), or a change in voicing diplophonia (Di) [8].

2. METHOD

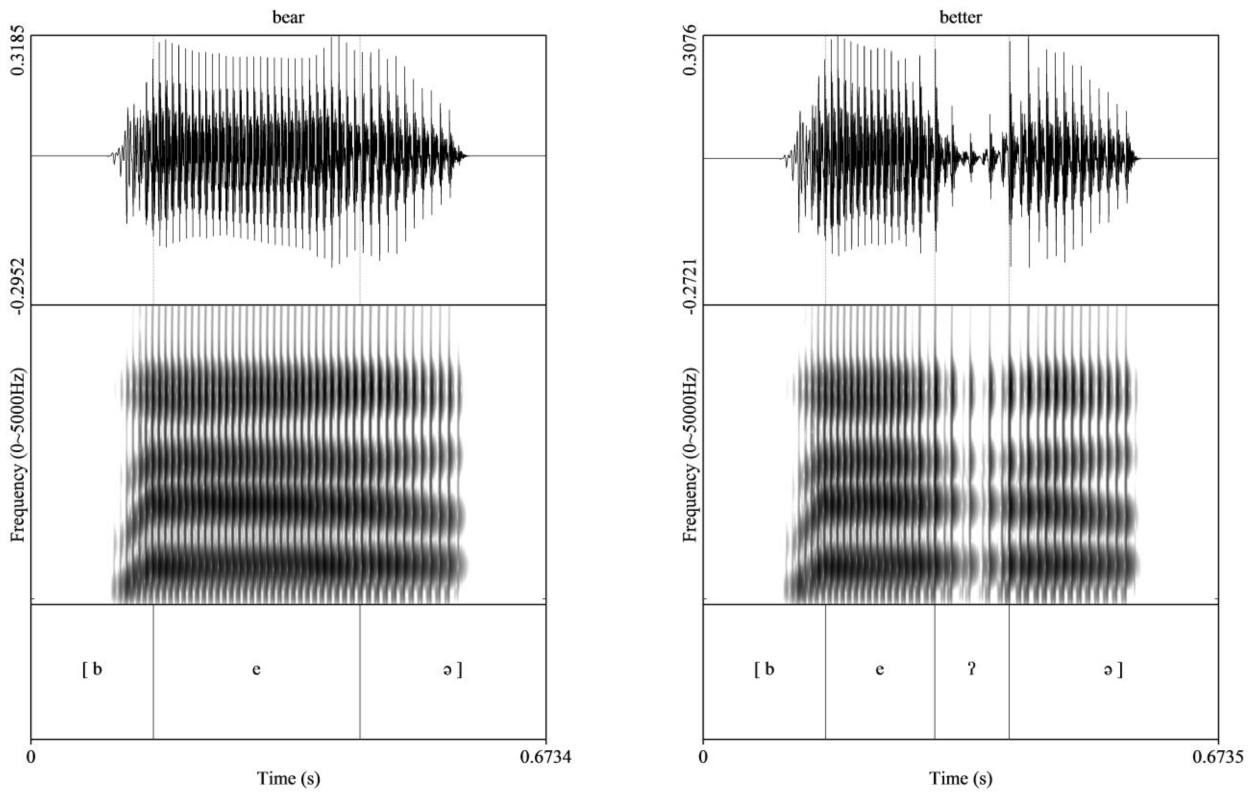
2.1. Subjects

A total of 27 subjects were tested: 15 native English speakers from London (6 males, 9 females) and 12 native Japanese speakers (5 males, 7 females). The native English speakers were 20-33 years old (median 23 years) and have spent most of their lives in London. The Japanese subjects were 22-38 years old (median 33 years), and had lived in London for up to 6 months (median 2.4 months). None of the subjects had any diagnosed hearing or speech impairments.

2.2. Stimuli and apparatus

The extreme stimuli ‘bear’ and ‘better’ were constructed from a natural model using the Klatt synthesizer [11, 12]. The timing, pitch, formant frequencies and voicing amplitude for ‘bear’ were obtained from an isolated utterance recorded by a native British English speaker. To obtain a typical instance of ‘better’ that included a glottal stop, a speaker was recorded reading a story and subsequently discussing its contents in an interview. A prototypical instance of ‘better’ as [beʔə] was then extracted from the interview task. The extreme ‘better’ [beʔə] stimulus was then generated by manipulating the synthesis of ‘bear’ according to the prototype. The key differences were that the ‘better’ extreme had a fundamental frequency (F0) dip of 35Hz; an amplitude of voicing (Ampl) dip of 14dB; and a diplophonia (Di) increase of 70 compared to ‘bear’. The value for Di shows a percentage increase or decrease in the duration of the closed phase of alternate voicing cycles. See Fig. 1 for waveforms and spectrograms of the extremes. The exact values used were chosen to ensure that the phoneme boundary for English listeners would occur around the middle of the continuum.

Figure1: Synthetic extremes ‘bear’ [beə] and ‘better’ [beʔə].



Seven different continua were then constructed between the extremes:

1. Fundamental frequency only (F0)
2. Voicing amplitude only (Ampl)
3. Diplophonia only (Di)
4. Fundamental frequency and voicing amplitude (F0Ampl)
5. Fundamental frequency and diplophonia (F0Di)
6. Voicing amplitude and diplophonia (AmplDi)
7. All three cues (F0AmplDi)

Each continuum was divided into 8 steps: F0 decreased by 5 Hz each step, Ampl decreased by 2 dB each step, and Di increased by 10 at each step. Thus, there were 56 stimuli in total, and each was presented 5 times in the listening experiment making a total of 280 stimuli heard by each participant.

2.3. Procedure

In order to familiarize the subjects with the stimuli, they were played the extreme token of ‘bear’ (with F0 100 Hz, Ampl 55 dB and Di 0 for /e/), and the extreme token of ‘better’ (with F0 65 Hz, Ampl 41 dB and Di 70 for /e/) several times. This familiarization was repeated during a listening test break that was given after stimulus 168. This was done to reduce any drift in perceptual boundaries over the course of the experiment.

The listening task was conducted in a sound-treated booth, with the subjects listening through Sennheiser 280 headphones connected to a laptop computer (Dell Inspiron) at a comfortable listening level.

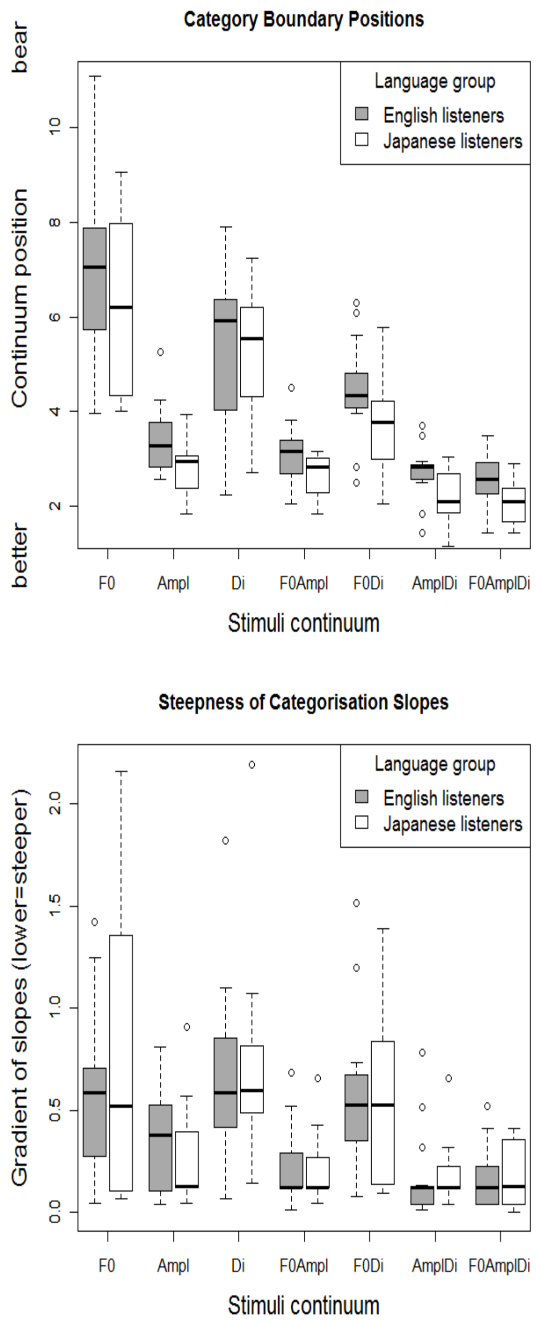
The category boundary position and steepness of categorisation slopes were estimated for each continuum for each subject by fitting a psychometric function to their labelling results. The fitting was done with the PSigniFit toolbox for MATLAB [4]. The effect of listener group on the category boundary position and steepness of the slopes along each continuum were then investigated using listener as a random variable.

Independent sample t-tests (two-tailed) were used to test the significance of the shift in mean category boundary and in mean steepness of the slopes between English and Japanese listeners for each stimulus condition. A significance level of $p = 0.05$ was used.

3. RESULTS

Both language groups were able to use the three acoustic cues singly and in combination to consistently tell apart ‘bear’ and ‘better’ (see Fig. 2). As might have been expected, there is less variability in category boundary position and the steepness of categorisation slopes across speakers for the multiple-cue continua than for the single cue continua.

Figure 2: Mean category boundary positions (left) and mean steepness of the categorisation function (right) for the seven conditions for the two listener groups.



In addition, as the number of cues increased across conditions, the category boundary location moved towards the 'bear' extreme and the gradient of slopes became steeper. These are because the quantity of acoustic change at each stimulus step is greater for the multiple-cue conditions than for the single cue conditions.

Although the overall pattern was similar across the language groups, significant differences did occur. Independent sample t-test of category boundary means across groups shows that there was a significant shift to an earlier boundary by the Japanese listeners in the AmplDi continuum (shift = -0.51step, p = 0.029) and in the F0AmplDi continuum (shift = -0.45step, p = 0.034). Since the largest shifts in boundary seem to be in those continua that include the Ampl cue, a possible interpretation is that the Japanese listeners are more sensitive to amplitude of voicing changes than English listeners.

There was no significant difference in the steepness of any categorisation slopes between two language groups.

4. DISCUSSION

The results indicate that both English and Japanese listeners can use changes in F0, Ampl and Di to make phonetic labelling distinctions. This may have been expected since both languages do use glottal stops, although not in the same phonetic environments. However, there seems to be a difference in the sensitivity of the two groups to the amount of phonetic change required to signal a contrast, with the Japanese listeners being more sensitive than the English listeners, choosing the glottal stop interpretation with significantly less acoustic change in the stimuli. One possible interpretation of this result is that the English listeners treated this task as a phonological choice, and were willing to accommodate sub-phonemic variation within the two phonological categories; while the Japanese listeners treated this task more as an auditory choice, so positioned the category boundary at the earliest place where sufficient evidence had built up to conclude that the stimulus was auditorily different to 'bear'.

It would be interesting to extend this study to include other languages which use the glottal stop as a phoneme, for example Hawaiian. If this difference in interpretation of the objective of the task is confirmed by further experiments, this result could also inform models of phonological category acquisition which distinguish between non-native phones which match or do not match native language categories [2].

5. REFERENCES

[1] Akamatsu, T. 2007. The phonological status of the glottal stop in British English. *La Linguistique* 43(2), 3-18.

[2] Best, C.T., McRoberts, G.W., Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J. Acoust. Soc. Am.* 109(2), 775-794.

[3] Bloch, B. 1950. Studies in colloquial Japanese VI phonemics. *Language* 26(1), 86-125.

[4] Bootstrap-software. *http://www.bootstrap-software.org/psignifit/*

[5] Collins, B., Mees, I.M. 1996. Spreading everywhere? How recent a phenomenon is glottalisation in received pronunciation? *English World-Wide* 17(2), 175-187.

[6] Cruttenden, A. 2008. *Gimson's Pronunciation of English: Seventh Edition*. Hachette: Hodder Education.

[7] Fabricius, A.H. 2000. *T-glottalling between Stigma and Prestige: A Sociolinguistic Study of Modern RP*. Ph.D. thesis, Copenhagen Business School.

[8] Hillenbrand, J.M., Houde, R.A. 1996. Role of F0 and amplitude in the perception of intervocalic glottal stops. *Journal of Speech and Hearing Research* 39, 1182-1190.

[9] Ito, K., Strange, W. 2009. Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *J. Acoust. Soc. Am.* 125(4), 2348-2360.

[10] Iverson, P., Kuhl P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, 47-57.

[11] Klatt, D.H. 1980. Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.* 67(3), 971-995.

[12] Klatt. D.H., Klatt, L.C. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 87 (2), 820-857.

[13] Lisker, L., Abramson, A.S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20 (3), 384-422.

[14] Lisker, L., Abramson, A.S. 1970. The voicing dimension: some experiments in comparative phonetics. *Proc. 6th ICPhS* Prague, 563-567.

[15] Vance, T.J. 1987. *An Introduction to Japanese Phonology*. Albany: State University of New York Press.

[16] Wells, J.C. 1982. *Accents of English 1: An Introduction*. Cambridge: Cambridge University Express.

[17] Wenck, G. 1966. *The Phonemics of Japanese: Questions and Attempts*. Harrassowitz: Wiesbaden.

[18] Yamaguchi, T. 2007. *Japanese Linguistics: An Introduction*. London: Continuum.